

„Intention to treat“ oder „treatment as received“ – Umgang mit Abbrechern in der Forschung zur Straftäterbehandlung

Methodische Überlegungen und Beispiele aktueller Studien

Johann Endres · Maike M. Breuer · Mark Stemmler

Eingegangen: 7. September 2015 / Angenommen: 23. November 2015 / Online publiziert: 14. Dezember 2015
© Springer-Verlag Berlin Heidelberg 2015

Zusammenfassung Viele Studien zur Straftäterbehandlung sind dem Vorwurf fehlender internaler Validität ausgesetzt, denn experimentelle Studien mit randomisierter Zuweisung sind schwer zu realisieren, und die Vergleichsgruppen in quasiexperimentellen Studien sind meist nicht völlig äquivalent. Deswegen werden häufig spezielle Matching-Prozeduren angewandt, oder ein potenzieller „selection bias“ durch Einflüsse von Störvariablen wird kontrolliert, indem diese als Kovariaten behandelt werden. Ein weiteres Problem ergibt sich jedoch durch hohe Abbruchraten, denn die Merkmale, die zum Behandlungsabbruch führen, sind häufig zugleich Risikofaktoren für einen kriminellen Rückfall. Es wird argumentiert, dass das Weglassen der Abbrecher in einem „Treatment-as-received“(TR)-Ansatz dazu führen kann, dass echte Behandlungseffekte durch Stichprobeneffekte überlagert werden. Erhöhte Rückfallraten bei Abbrechern können auf negative Wirkungen des Abbruchs selbst (z. B. Stigmatisierung) zurückgehen, sind aber wohl häufiger darauf zurückzuführen, dass Abbrecher von vornherein ein höheres Rückfallrisiko aufweisen. Am Beispiel einiger neuerer Studien wird gezeigt, dass scheinbare Behandlungseffekte oft verschwinden, wenn der adäquatere „Intention-to-treat“(ITT)-Ansatz verwendet wird, bei dem die Abbrecher weiterhin der Behandlung zugerechnet werden.

Schlüsselwörter Straftäterbehandlung · Abbrecher · Behandlungsforschung · Rückfall

Intention-to-treat and treatment-as-received: dealing with dropouts in research on offender treatment

Methodological considerations and examples of recent studies

Abstract Lack of internal validity is a frequent problem in evaluation studies of offender treatment, as randomized controlled trials are difficult to implement and comparison groups in quasi-experimental studies are often not completely equivalent. Special matching procedures or statistical control of selection bias by using potentially interfering variables as covariates can provide a solution to some of these problems; however, a high dropout rate from offender treatment programs presents another problem, as the variables that cause attrition are often risk factors for reoffending. It is argued that a treatment-as-received (TR) approach which compares treatment completers with the comparison group may confound genuine treatment effects with selection effects. Particularly high recidivism rates in dropouts may be due to negative effects of attrition (e.g. stigmatization) but may more likely reflect a higher a priori risk for reoffending. Using data from some recent treatment studies, the authors show that alleged positive effects often disappear when a more adequate intention-to-treat (ITT) approach is used in which dropouts are retained in the treatment sample.

Keywords Offender treatment · Attrition · Therapeutic research · Recidivism

Dr. J. Endres (✉) · Dr. M. M. Breuer
Kriminologischer Dienst des bayerischen Justizvollzugs,
Schuhstr. 41,
91052 Erlangen, Deutschland
E-Mail: johann.endres@jva-er.bayern.de

Prof. Dr. M. Stemmler
Institut für Psychologie, FAU Erlangen,
Nägelsbachstr. 49c,
91052 Erlangen, Deutschland

Die Behandlungsforschung will empirische Erkenntnisse darüber gewinnen, welche Effekte die Teilnahme an einer Intervention (Therapie, Training, Behandlung) hinsichtlich relevanter Zielgrößen hat. Bei der Straftäterbehandlung ist dies häufig die kriminelle Rückfälligkeit innerhalb eines festgelegten Zeitraums. Alternativ können aber auch andere Maße, insbesondere Veränderungen bei dynamischen Risikofaktoren (z. B. soziale Kompetenzen, Aggressionsneigung oder deliktbegünstigende Einstellungen), untersucht werden.

In der Praxis lautet die Fragestellung häufig: Was „bringt“ die Behandlung? Was wäre anders, wenn man den Teilnehmern diese Behandlung nicht anbieten würde, sondern entweder keine oder eine Standardbehandlung (z. B. „Normalvollzug“)? Diese im Irrealis stehenden Fragen sind jedoch so nicht beantwortbar, weil eben eine Person nur entweder an einer Behandlung teilnehmen oder nicht teilnehmen kann. Der bloße Vergleich von Parametern vor und nach einer Intervention (Ein-Gruppen-Versuchsplan) ist wenig aussagekräftig, weil sich hier Behandlungseffekte nicht von anderen Einflüssen, z. B. Reifung, Gewöhnung und sonstigen Lernprozessen unterscheiden lassen; darin liegt das fundamentale Problem der kausalen Inferenz [24]. Man greift deshalb auf experimentelle oder quasiexperimentelle Untersuchungsansätze zurück, bei denen man jeweils behandelte Personen mit Unbehandelten vergleicht. Diese Ansätze haben jedoch ihrerseits ihre methodischen Tücken, die die Validität (Gültigkeit) der so gewonnenen Erkenntnisse gefährden können. Nachfolgend wird anhand von einigen neueren Untersuchungen und theoretischen Überlegungen das methodische Problem des Umgangs mit Behandlungsabbrechern und -verweigerern vertiefend diskutiert. Es wird argumentiert, dass das Herausnehmen der Abbrecher aus dem Vergleich von Behandelten und Unbehandelten, wie es in manchen Evaluationsuntersuchungen geschieht („treatment as received“, TR), zu einer Fehleinschätzung der Behandlungseffekte führt. Wenn die Abbrecher schlechtere Ergebnisse erzielen als Unbehandelte, führt das zu einer Überschätzung des Effekts der Behandlung. Aus diesem Grund sollten die Abbrecher entsprechend der ursprünglichen Zuweisung in der Behandlungsbedingung verbleiben („intention to treat“, ITT).¹

¹Obwohl es für die Darstellung und Interpretation der Ergebnisse erhebliche Unterschiede macht, ob ein ITT- oder ein TR-Ansatz verfolgt wird, spielt dieses Thema in der Literatur zu Forschungsmethodik und Evaluation kaum eine Rolle. In den gängigen deutschsprachigen Lehrbüchern [2] oder Handbüchern [10] kommt es überhaupt nicht vor. In medizinischen Lehrbüchern zu Forschungsmethoden (z. B. [6], S. 58) hingegen wird meist ein ITT-Ansatz empfohlen.

Experimentelle und quasiexperimentelle Untersuchungspläne

Um die Frage zu beantworten, wie eine Behandlung wirkt, reicht es nicht aus, nur den Zustand vorher und nachher zu vergleichen. Dabei zu findende Unterschiede können ja nicht eindeutig der Behandlung zugeordnet werden, sondern könnten auch aus sonstigen Einflüssen und Prozessen (z. B. Gewöhnung an eine Umgebung, normale Reifungsvorgänge) resultieren. Es kommt vielmehr darauf an, die Veränderung durch die Behandlung mit den Veränderungen in einer Kontrollgruppe (ohne diese Behandlung, aber ansonsten äquivalent, d. h. im Idealfall mit identischen Ausprägungen hinsichtlich wesentlicher Merkmale; Zustand „ceteris paribus“; [24]) in Beziehung zu setzen. Wenn z. B. in einer Stichprobe von Personen, die an einem Anti-Gewalt-Training teilgenommen haben, die Häufigkeit aggressiver Verhaltensweisen deutlich stärker zurückgeht als bei Personen, die nicht teilgenommen haben, kann dieser Effekt kausal auf die Maßnahme zurückgeführt werden – insoweit die beiden Personengruppen sich nicht durch andere Merkmale unterscheiden, die den Effekt erklären könnten.

Es gibt zwei methodische Probleme, die den Rückschluss von derartigen Effekten auf eine bestimmte Behandlung gefährden können. Das eine sind Zufallseinflüsse, die bei fast allen natürlichen oder sozialen Vorgängen zu erwarten sind. Sie in den Griff zu bekommen, ist Aufgabe der inferenzstatistischen Prüfung. Etwas vereinfacht gesagt, wird so vorgegangen, dass nur Effekte als „statistisch signifikant“ akzeptiert werden, die mit hoher Wahrscheinlichkeit nicht auf Zufallsschwankungen in den betrachteten Merkmalen bzw. auf zufällige Abweichungen bei der Ziehung einer Stichprobe aus einer Grundgesamtheit zurückgeführt werden können.

Das zweite methodische Problem liegt in der zu fordernden Äquivalenz zwischen Behandlungs- und Kontrollgruppe. Wenn sich die beiden Gruppen in relevanten Merkmalen deutlich unterscheiden, wird der Schluss von Effekten auf die Maßnahme zweifelhaft. Waren z. B. die Teilnehmer an einem Anti-Gewalt-Training zuvor schon deutlich weniger aggressiv als die Nichtteilnehmer (etwa weil die Hochaggressiven das Training ablehnten und die Teilnahme verweigerten, oder weil die Trainer sie als schwierig ablehnten und nur gut angepasste Bewerber auswählten), kann die anschließend größere Friedfertigkeit der Teilnehmer keinesfalls kausal auf das Training zurückgeführt werden; sie bestand ja schon zuvor. Aber auch wenn umgekehrt die Teilnehmer am Anfang deutlich aggressiver waren als die Nichtteilnehmer, ist der Schluss problematisch: So könnte der Rückgang aggressiver Verhaltensweisen in der Behandlungsgruppe allein auf die zunehmende Vertrautheit der Personen untereinander zurückzuführen sein, während in der Kontrollgruppe (die z. B. eine unспе-

zifische Maßnahme erhält, etwa eine Gruppendiskussion) die Veränderung auch deshalb viel geringer ausfallen kann, weil hier die Ausgangsrate schon sehr niedrig war („Boden-Effekt“). Diese Stichprobenunterschiede führen zu einem „Bias“ (Verzerrung, Störeffekt), der den Rückschluss von der Veränderung auf die Maßnahme infrage stellt.

Der methodische „Goldstandard“, um die Äquivalenz zwischen Behandlungs- und Kontrollgruppe sicherzustellen, wird in einer experimentellen Vorgehensweise mit kontrollierter randomisierter Aufteilung der Personen auf Behandlungs- und Kontrollgruppe gesehen: Die Entscheidung, wer behandelt wird und wer nicht, treffen die Forscher und verwenden dafür ein Zufallsverfahren. Dies kann beispielsweise dadurch geschehen, dass man innerhalb der Grundgesamtheit von Personen, die für die Behandlung in Betracht kommen, Paare von sehr ähnlichen Personen bildet (Matching) und dann jeweils auslost, welche der beiden Personen in die Behandlungs- oder in die Kontrollgruppe kommt. Dies garantiert, dass sich die beiden Gruppen hinsichtlich der Merkmale, die beim Matching verwendet werden, allenfalls durch zufallsbedingte Schwankungen unterscheiden können. Dies gilt allerdings nur für den Zeitpunkt der Aufteilung; wegen selektiver, nichtzufälliger Ausfälle kommt es in der Folge dann doch häufig zu Unterschieden zwischen Behandlungs- und Kontrollgruppe (s. nächster Abschnitt „Behandlungsabbruch als verzerrender Effekt („bias“)“).

Aus naheliegenden, insbesondere rechtlichen, ethischen und politischen Gründen sind echte experimentelle Versuchspläne in der forensischen Behandlungsforschung sehr selten (zu den wenigen Ausnahmen gehören [18] und [16]). Sie lassen sich nur dann realisieren, wenn es keine verbindliche gesetzliche Pflicht zur Behandlung und keinen entsprechenden Rechtsanspruch gibt, wenn die verfügbaren Behandlungsplätze nicht für alle Behandlungsbedürftigen und -bereiten ausreichen oder wenn das Vorenthalten der Behandlung nicht als eine gravierende Benachteiligung angesehen werden kann. Die weitere unabdingbare Voraussetzung ist natürlich auch, dass die für die jeweiligen Entscheidungen Verantwortlichen bereit sind, sich dem Zufallsprinzip zu unterwerfen.

In der Forschungspraxis sehr viel häufiger sind quasi-experimentelle Untersuchungspläne, bei denen eine vorgefundene Aufteilung in Behandelte und Unbehandelte zugrunde gelegt wird. Diese vorfindliche Aufteilung kann sich aus juristischen Festlegungen ergeben (z. B. richterliche Entscheidung für eine bestimmte Sanktion, die Behandlungsoptionen einschließt, wie die Unterbringung in einer Entziehungsanstalt nach § 64 StGB), aus regionalen Unterschieden (z. B. Verfügbarkeit einer Behandlung nur in einigen Bundesländern oder Einrichtungen) oder aus Entscheidungen der potenziellen Teilnehmer (z. B. Bewerbung als Aufnahmevoraussetzung). In allen diesen Fällen ist nicht gewährleistet, dass die Behandlungsgruppe und

die Vergleichsgruppe hinsichtlich wesentlicher Merkmale äquivalent sind. Vielmehr können Stichprobenverzerrungen („biases“) nicht ausgeschlossen werden; das bedeutet, dass Merkmale, die die Auswahl für die Behandlung beeinflussen, auch Auswirkungen auf die Zielgrößen haben. Wenn beispielsweise Personen mit guter Schulbildung bevorzugt in ein Anti-Gewalt-Training aufgenommen werden, können scheinbare Behandlungseffekte (bessere soziale Integration) nicht eindeutig dem Training zugeordnet werden, da sie möglicherweise aus den günstigeren schulischen Qualifikationen (und damit vermutlich größeren Chancen auf dem Arbeitsmarkt) resultieren. Ob solche Selektionseffekte vorliegen, kann aber nachträglich überprüft werden (z. B. Vergleiche hinsichtlich Altersstruktur, Vorstrafenbelastung, Deliktarten). Soweit diese Vergleiche keine ausreichende Äquivalenz zeigen, können bestehende Unterschiede bei der Auswertung berücksichtigt und in ihren Auswirkungen statistisch kontrolliert werden, indem zugleich mit dem Einfluss der Behandlung auch der Einfluss der weiteren Variablen analysiert wird (technisch gesprochen, indem die möglichen konfundierenden oder störenden Merkmale als Kovariaten modelliert werden). Allerdings setzt dieses Vorgehen voraus, dass die entsprechenden Störeinflüsse bekannt sind und zuvor erfasst werden können. Es lässt sich nie völlig ausschließen, dass weitere Störvariablen wirksam waren, an die man nicht gedacht hat oder die man nicht zuverlässig messen kann [21].

Behandlungsabbruch als verzerrender Effekt („bias“)

Ein besonderer methodischer Störeffekt geht daraus hervor, dass Behandlungen aus unterschiedlichen Gründen (durch die Teilnehmer selbst, aber auch durch die Institution) abgebrochen oder verweigert werden können. Wie es durch selektiven Ausfall dazu kommen kann, dass sich Behandlungs- und Kontrollgruppe deutlich unterscheiden, auch wenn sie nach der ursprünglichen Zuteilung noch völlig äquivalent waren, illustriert Abb. 1 (für den vereinfachten Fall, dass die Variation innerhalb einer Population nur aus vier unterschiedlichen Merkmalsausprägungen besteht).

Abbrecherquoten belaufen sich bei der Straftäterbehandlung im Durchschnitt auf 27,1% (Metaanalyse von [17]; 96 Studien gingen ein); der Anteil war höher bei unterschiedlichen Arten von ambulanten Behandlungen als bei stationären, am höchsten jedoch (57,7%) bei Studien mit jugendlichen Insassen. Abbrecher sind häufig jünger, haben mehr Vorstrafen und häufiger Diagnosen einer antisozialen Persönlichkeitsstörung [17, 3]. Deshalb liegt die Vermutung nahe, dass Merkmale wie Selbstbeherrschung, Durchhaltevermögen, Frustrationstoleranz und Änderungsmotivation das Abbrechen beeinflussen. In vielen Untersuchungen haben Abbrecher höhere Rückfallquoten als diejenigen Per-

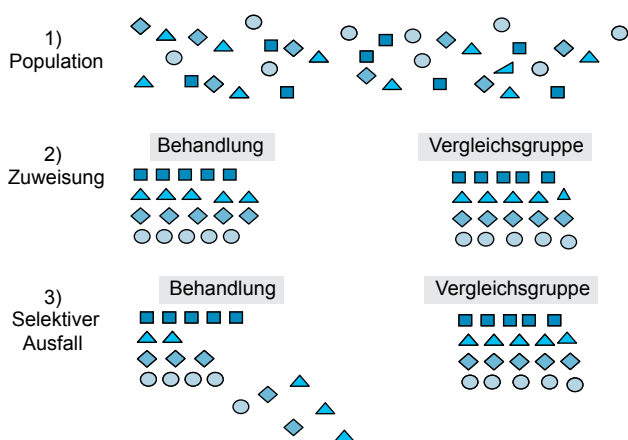


Abb. 1 In der Population sind verschiedene Merkmalsausprägungen vorhanden (Schritt 1). Durch die experimentelle Zuweisung wird die Äquivalenz zwischen der Behandlungsbedingung und der Vergleichsgruppe hergestellt (Schritt 2). Selektive Ausfälle gefährden diese Äquivalenz (Schritt 3; Ausfalleffekt)

sonen, die eine Behandlung regulär abschließen. Behandlungsabbruch ist also ein Merkmal, das häufig sowohl mit der Zusammensetzung der beiden zu vergleichenden Gruppen als auch mit den Zielvariablen in Beziehung steht und somit die Gültigkeit von Rückschlüssen aus den Daten auf einen Effekt der Behandlung infrage stellen kann.

Wenn man die Wirkung einer Behandlung evaluieren will, muss man demnach zwischen unterschiedlichen substantziellen und methodischen Effekten unterscheiden:

1. Im Zentrum des Interesses steht *der intendierte kausale Effekt der Behandlung*, ihre *Wirksamkeit*. Straftäterbehandlung soll kriminalpräventiv wirken, d. h., bei den Zielgruppen, für die sie angeboten wird, die Rückfälligkeit senken. Wenn 100 Personen einer Behandlung zugeteilt werden, sollte ihre Rückfallrate niedriger ausfallen als bei 100 vergleichbaren Personen, denen diese Behandlung nicht angeboten wird. Dieser Effekt ist ausschlaggebend für die Entscheidungen hinsichtlich der Zuteilung von Ressourcen oder auch für Empfehlungen, eine Behandlung anzutreten. Diese Forderung nach einem positiven Saldo gilt natürlich für die gesamte Zielgruppe, die für die Behandlung vorgesehen ist, unabhängig davon, zu welchem Anteil sie zu der Behandlung antreten und diese auch vollständig absolvieren.
2. Davon zu unterscheiden ist die *Leistungsfähigkeit* der Behandlung im Einzelfall. Diese zeigt sich in einem *Dosis- oder Abschlusseffekt*, d. h., die Behandlung hat Wirkungen in Abhängigkeit vom Ausmaß der Teilnahme oder nur bei vollständiger Teilnahme. Bei Abbrechern wäre dann entweder kein Effekt oder jedenfalls ein deutlich geringerer Effekt zu erwarten. Wenn ein Dosis-Effekt wirksam ist, wäre zu erwarten, dass der Abstand zwischen Abbrechern und Absolventen vom

Zeitpunkt des Abbruchs abhängt und bei frühem Abbruch deutlich größer ausfällt als bei spätem Abbruch.

3. Möglicherweise gibt es aber auch *nichtintendierte negative Effekte einer Behandlung* [15], auch wenn dies in der bisherigen Literatur zur Straftäterbehandlung selten thematisiert worden ist [4]. Ähnlich wie bei Medikamenten können auch bei Trainings oder psychotherapeutischen Interventionen „Risiken und Nebenwirkungen“ nicht ausgeschlossen werden (z. B. negative Lernvorgänge in Gruppen, psychische Labilisierung durch harte Konfrontationen, Stigmatisierung durch Verlegung in eine Sexualtäterabteilung). Derartige negative Effekte könnten bei manchen Verfahren möglicherweise durchgängig auftreten; näher liegt aber die Vermutung, dass sie sich evtl. auf einzelne Zielgruppen beschränken.
4. *Stichprobeneffekte* können durch eine entweder von vorneherein selektive Auswahl der Behandelten (*Selektionseffekt*) oder später durch selektiven Drop-out (*Ausfalleffekt*) zustande kommen. Sie sind rein methodischer Art, d. h., die Unterschiede in den Ergebnissen sind nicht auf die Prozesse während und aufgrund der Behandlung zurückzuführen, sondern bestehen unabhängig davon, aufgrund unterschiedlicher Behandlungsvoraussetzungen. Möglicherweise sind die Abbrecher diejenigen Personen aus der Behandlungsgruppe, die sowieso die schlechteren Risiken darstellten, ungünstigere Voraussetzungen hatten und auch ohne Behandlung, oder wenn der Abbruch verhindert worden wäre, die schlechteren Ergebnisse gehabt hätten. Das erscheint naheliegend bei Abbrüchen aus disziplinarischen Gründen (jemand wird wegen Fehlverhaltens aus der Maßnahme ausgeschlossen). Aber auch mangelnde Motivation kann ein Indikator für eine ungünstigere Ausgangssituation sein: Jemand, der eine Behandlung nicht weiter fortsetzt, weil er merkt, dass er nicht davon profitiert, könnte auch aus genau diesem Grund ein höheres Risiko haben. Ausfalleffekte können aber manchmal auch in positiver Richtung wirken (wenn z. B. Personen abbrechen, bei denen sich schon zu Beginn der Behandlung eine Besserung einstellt und die deshalb die Behandlung nicht mehr nötig haben und nicht mehr von ihr profitieren können).
5. Ein echter inhaltlicher *Abbrucheffekt* resultiert demgegenüber daraus, dass das Ereignis Behandlungsabbruch selbst eine negative Wirkung ausübt. Dies kann beispielsweise durch Stigmatisierungsvorgänge (der Abbrecher wird als Problemfall definiert und diskriminiert) oder durch motivationale Auswirkungen (der Abbrecher zweifelt an sich und gibt Bemühungen zur Veränderung auf) erfolgen. Beispielsweise erhalten Strafgefangene, die eine Behandlung abbrechen, in der Folge häufiger keine Vollzugslockerungen und werden unzureichend auf die Entlassung vorbereitet [3]. Oder die Erfahrung, aus der Behandlungsmaßnahme entfernt

zu werden, führt zum Verlust der Hoffnung auf erreichbare Veränderungen. Personen, deren Behandlung abgebrochen wurde, werden aufgrund derartiger Prozesse möglicherweise häufiger rückfällig, als wenn sie erst gar nicht die Behandlung begonnen hätten.

In der Gesamtbilanz einer Behandlung schlagen sich also neben ihrer Wirksamkeit (Effekt Nr. 1) und ihrer Leistungsfähigkeit im Einzelfall (Nr. 2) auch ein evtl. schädlicher Behandlungseffekt (Nr. 3) und der Abbrucheffekt (Nr. 5) jeweils in negativer Richtung nieder. Beim Ausfalleffekt (Nr. 4) hingegen handelt es sich um ein rein statistisches Phänomen, das unabhängig von der Wirksamkeit einer Behandlung auftreten kann und diese seinerseits nicht beeinflusst.

Aus einem Vergleich der Ergebnisse zwischen Behandlungs- und Kontrollgruppen ist in der Regel nicht zu erkennen, worauf mögliche Unterschiede beruhen. Das Ziel der Evaluation besteht jedoch darin, die eigentlichen Behandlungseffekte zu ermitteln (Nr. 1, daneben auch Nr. 2 und Nr. 3). Der Abbrucheffekt (Nr. 5) kann, wenn er auftritt, Anlass geben, die Programmvorgabe oder die Kriterien der Auswahl und des Ausschlusses von Teilnehmern zu modifizieren. Den Ausfalleffekt (Nr. 4) und andere Störeffekte gilt es durch geeignete methodische Vorkehrungen bei der Auswertung zu eliminieren. Ein entscheidender Gesichtspunkt ist dabei, wie mit den Behandlungsabbrechern methodisch verfahren wird.

Intention-to-treat- und Treatment-as-received-Analysen

In Behandlungsstudien gibt es zwei unterschiedliche Ansätze, wie mit Behandlungsabbrechern umgegangen werden kann [7, 13, 22]:

- Nach dem „Intention-to-treat“-Ansatz (auch „intent-to-treat“ genannt; nachfolgend abgekürzt ITT) bleiben alle Personen, bei denen eine Behandlung geplant war, also auch die Abbrecher, und die Personen, die gar nicht erst angetreten sind, in der Behandlungsgruppe, deren Ergebnisse mit denen der Kontrollgruppe verglichen werden. Die dahinter stehende Überlegung ist die, dass die anfänglich definierte Einteilung in eine Treatment- und eine Kontrollbedingung im weiteren Verlauf für die Analysen maßgeblich sein sollte. Die vorgesehene Behandlung sollte insgesamt wirken, und nicht nur bei denen, die bis zum Ende teilnehmen. Vielmehr sollte die Gesamtgruppe der Personen, die für die Behandlung vorgesehen sind, bessere Ergebnisse aufweisen als die Gesamtgruppe der Personen, denen die Behandlung nicht zuteilwird. Der Vergleich kann allerdings nicht die optimale Leistungsfähigkeit einer Behandlung sichtbar

machen, weil auch Personen als behandelt klassifiziert werden, die nur kurzzeitig an ihr teilhatten.

- Hingegen werden nach dem alternativen „Treatment-as-received“-Ansatz (nachfolgend abgekürzt TR; auch „Per-protocol“-Analyse genannt) die Behandlungsabbrecher und -verweigerer aus dem Vergleich der Behandlungsgruppe mit der Kontrollgruppe herausgenommen. In der Behandlungsbedingung verbleiben nur die Personen, bei denen die vorgesehene Behandlung im vollen Umfang durchgeführt worden ist. Dies entspricht dem Standpunkt vieler Praktiker, die argumentieren, sie hätten ja nur eine Chance gehabt, diejenigen Personen positiv zu beeinflussen, die auch an der Behandlung teilgenommen haben, und die Misserfolge bei Abbrechern nicht der Behandlung zurechnen lassen wollen. Der Nachteil dieses Auswertungsansatzes liegt jedoch darin, dass selektive Abbrucheffekte häufig die Behandlung scheinbar begünstigen, ohne wirklich einen Behandlungseffekt zu belegen. Wenn Abbrecher (wie häufig der Fall) die schlechteren Risiken aufweisen und diese schlechteren Risiken zu häufigeren Rückfällen führen, verbleiben in der Behandlungsgruppe mehr Personen, die von vorneherein eine günstigere Prognose haben als die Kontrollgruppe.

Für Zwecke der Wirksamkeitsforschung erscheint es zwingend, dem ITT-Ansatz zu folgen. Nur dann lässt sich überzeugend darlegen, dass positive Effekte tatsächlich der Überlegenheit der Behandlung und nicht etwa dem Selektionseffekt (Ausscheiden negativer Risiken) zugeordnet werden können. Zusätzliche Vergleiche von Teilgruppen (Absolventen und Abbrecher) können Hinweise liefern, die für Fragen der Behandlungsindikation, die Vermeidung von Nebenwirkungen oder die Fortentwicklung der Behandlungsmethoden bedeutsam sind. Dies wird im Folgenden an einigen realen und fiktiven Beispielen erläutert.

Beispiel 1: Radiotherapie bei Brustkrebs

Bei der Behandlung von Brustkrebs nach Operationen wird in einer Metaanalyse [14] für die Behandlung mit Radiotherapie („Bestrahlung“) eine Mortalitätsrate von 21,4% innerhalb von 15 Jahren berichtet, im Vergleich zu 25,2% ohne Radiotherapie. Diese Reduktion des Sterblichkeitsrisikos um etwa ein Sechstel erscheint also sehr geeignet, diese Behandlung zu empfehlen. So weit die empirischen Erkenntnisse aus der genannten Metaanalyse. Die nachfolgenden Überlegungen sind fiktiver Art: Sollte sich nun aber herausstellen, dass die Hälfte der Patientinnen die Radiotherapie wegen zu starker Nebenwirkungen oder aus anderen Gründen abbricht und dass bei den Abbrecherinnen die Mortalitätsrate bei 30,0% liegt (und sich die Mortalitätsrate von 21,4% nur auf die regulär beendeten

Behandlungen ohne Abbrüche bezieht), würde das diese Empfehlung in Zweifel ziehen. Denn insgesamt würde sich dann durch die Entscheidung für die Radiotherapie das Mortalitätsrisiko sogar von 25,2% auf 25,6% leicht erhöhen ($21,4\% \cdot 0,50 + 30\% \cdot 0,50$).²

Beispiel 2 (fiktiv): Psychotherapie von Depression

Die für Beispiel 1 angestellten Überlegungen gelten analog auch für psychotherapeutische Behandlungen. Angenommen, von 200 Personen mit einer depressiven Störung werden ohne Behandlung nach einem halben Jahr 70 Personen (35%) symptomfrei oder gebessert sein, 70 Personen (35%) verschlechtert und 60 Personen (30%) unverändert (Abb. 2). Angenommen weiter, eine halbjährige Psychotherapie führt dazu, dass am Ende von 100 Teilnehmern, mit denen die Behandlung vollständig durchgeführt wurde, 60 verbessert sind, 10 verschlechtert und 30 unverändert. Dies scheint für die Therapie zu sprechen, weil der Anteil der Patienten, die eine Besserung aufweisen, deutlich über der Basisrate liegt.

Wenn man aber nun erfährt, dass es zusätzlich 100 Personen gab, die die Behandlung abgebrochen haben, von denen nur 10 eine Verbesserung aufwiesen, 30 keine Veränderung und 60 eine Verschlechterung, relativiert das die Erfolge.

Die Auswertung nach dem TR-Ansatz würde für einen erheblichen Behandlungseffekt sprechen (Verbesserung um 25 Prozentpunkte). Die Auswertung nach dem ITT-Ansatz spricht jedoch gegen eine Wirksamkeit der Behandlung: Einschließlich der Abbrecher ergeben sich identische Erfolgsquoten wie bei den Unbehandelten. Dabei wäre allerdings unklar, ob der Behandlungsabbruch zu der Verschlechterung führt oder ob umgekehrt die Personen mit schlechter Ansprechbarkeit für die Behandlung von vornherein die schlechtere Prognose hatten. Falls Ersteres zutrifft, wäre die Behandlung bei einem Teil der Probanden durchaus wirksam, hätte aber bei einer anderen Teilgruppe negative Effekte. Es würde dann darum gehen, Indikationen und Kontraindikationen zu erforschen und diagnostische Methoden zu finden, damit die Behandlung v. a. denjenigen angeboten wird, die auch von ihr profitieren. Falls Zweiteres zutrifft, müsste man die Behandlung wohl als insgesamt unwirksam beurteilen.

²Das gilt jedoch nur, falls das Scheitern der Radiotherapie nur vom Zufall abhängt und sich damit nicht vorhersagen lässt. Sollten aber Risikofaktoren für das Scheitern bekannt sein, und jemand gehört zu einer Gruppe, bei der mit 90%iger Wahrscheinlichkeit die Radiotherapie gut vertragen und nicht abgebrochen wird, stellt sich die Bilanz doch positiv dar ($21,4\% \cdot 0,90 + 30\% \cdot 0,10 = 22,25\%$). In diesem Fall wäre es doch rational, sich für die Radiotherapie zu entscheiden.

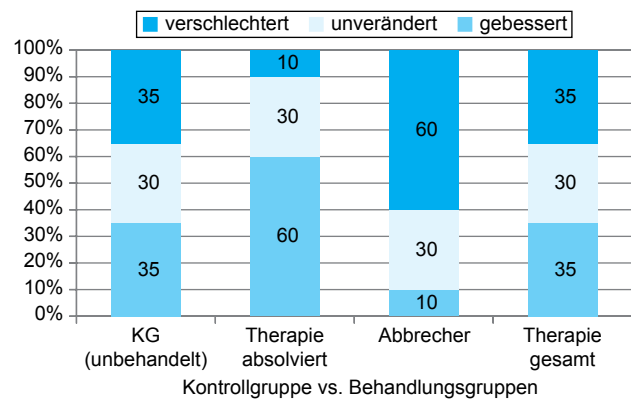


Abb. 2 Therapieergebnisse für eine unbehandelte Kontrollgruppe (KG), eine therapierte Stichprobe sowie separat für Absolventen und Abbrecher dieser Therapie. Die Abbildung illustriert, dass der Vergleich der Abbrecher mit der Kontrollgruppe einen Therapieerfolg vortäuscht, der tatsächlich nur in einem selektiven Ausfalleffekt besteht

Beispiel 3: Jugendstrafvollzug in freien Formen

In Baden-Württemberg gibt es zwei Projekte „Jugendstrafvollzug in freien Formen“: Zu Jugendstrafe ohne Bewährung verurteilte junge Männer werden in Projekten der Jugendhilfe in offenen Einrichtungen betreut [23]. Für die Teilnehmer dieser Projekte stehen, getrennt für Absolventen (59%) und für Abbrecher (41%) Rückfallzahlen für den Zeitraum drei Jahre nach der Entlassung zur Verfügung. Diese lassen sich vergleichen mit Informationen aus einer Totalerhebung zur Legalbewährung von aus dem geschlossenen Jugendstrafvollzug entlassenen Personen (Daten aus [12]; Sonderauswertung für Baden-Württemberg), ebenfalls mit Rückfallraten innerhalb von drei Jahren (Tab. 1):

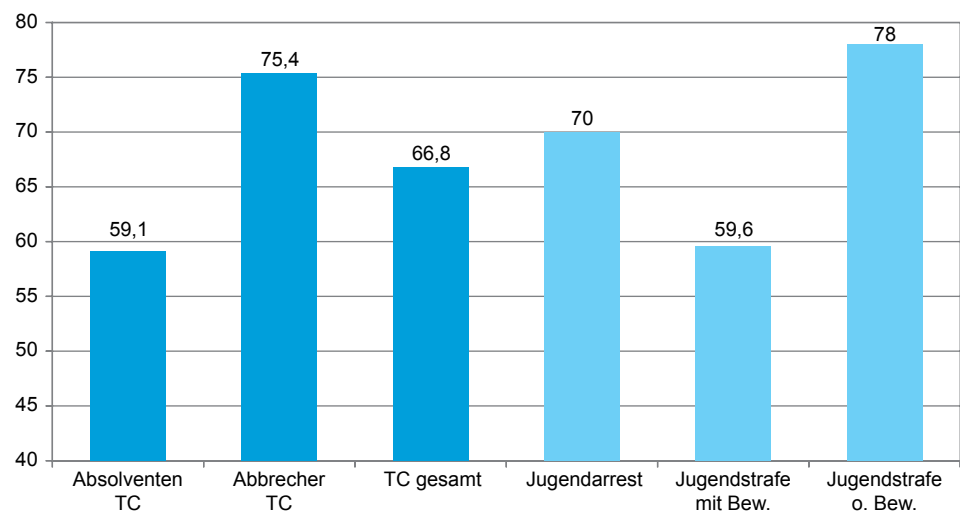
Wenn man (nach TR-Ansatz) die regulären Absolventen in Beziehung setzt zu den Vergleichsdaten aller aus dem Jugendstrafvollzug Entlassenen, ergibt sich eine leichte Überlegenheit des Projekts: Es sieht so aus, als hätte die Teilnahme an dem Projekt die allgemeine Rückfallquote von 72% auf 68% reduziert und die schweren Rückfälle von 36% auf 28% (also um fast ein Viertel).

Berücksichtigt man (nach dem ITT-Ansatz) allerdings bei der Treatment-Gruppe auch die Abbrecher („Rückverlegte“, 41%), die deutlich erhöhte Rückfallraten aufweisen (85% wurden mit irgendeiner Straftat rückfällig und 49% schwer), ergibt sich eine andere Bilanz: Von 100 jungen Strafgefangenen, die dem Modellprojekt zugewiesen wur-

Tab. 1 Rückfallraten (%) für Teilnehmer der „Projekt Chance“ und Vergleichsdaten

Rückfallart	Absolventen	Abbrecher	Vergleichsdaten
Jeder Rückfall	68	85	72
Schwerer Rückfall (erneute Inhaftierung)	28	49	36

Abb. 3 Rückfallraten (in %) für die Absolventen, die Abbrecher und die gesamten Teilnehmer des Trainingscamps (TC) sowie Vergleichsdaten für Rückfallraten nach Jugendarrest sowie nach Verurteilung zu Jugendstrafe mit Bewährung und Jugendstrafe ohne Bewährung



den, werden demnach 75 % mit irgendeiner Straftat rückfällig, und 36,6 % werden in schwerer Form rückfällig. Das stellt gegenüber der Vergleichspopulation jeweils sogar eine geringfügige Verschlechterung dar.³

Beispiel 4: Anti-Gewalt-Trainings in der österreichischen Bewährungshilfe

In der österreichischen Bewährungshilfe wird mit Probanden, die durch Gewaltdelikte aufgefallen sind, eine Variante des Anti-Gewalt-Trainings (AGT) durchgeführt, die sich stark auf konfrontative Methoden stützt („Heißer Stuhl“, [9]). Den 155 Absolventen dieser Trainings wurden durch Propensity Score Matching⁴ 155 Nichtteilnehmer zugeordnet (Kontrollgruppe). Die einschlägige Wiederverurteilungsquote betrug innerhalb eines Zeitraums von ca. 15 Monaten 24 % bei den Absolventen und ebenfalls 24 % bei den Nichtteilnehmern. Das Absolvieren des Trainings brachte also, so die Autoren, keinen nachweisbaren Vorteil für den weiteren Bewährungsverlauf, was vielleicht daran liegen könnte, dass auch die Nichtteilnehmer nicht völlig

³Es ist nicht auszuschließen, dass es trotzdem Behandlungseffekte in der Form gibt, dass die Rückfälle der Behandelten im Mittel etwas weniger schwerwiegend sind, oder dass es sonstige Erfolge gibt, z. B. bessere berufliche Qualifikationen, die sich nicht in der erhobenen Zielgröße abbilden. Hierzu gibt es aber keine Daten. Ein weiterer methodischer Problempunkt dieser Studie ist freilich darin zu sehen, dass nicht klar ist, nach welchen Kriterien die Teilnehmer dem Projekt zugeteilt wurden. Aufgrund der Rahmenbedingungen („freie Formen“) ist jedoch davon auszugehen, dass für die Teilnahme eher Verurteilte mit günstigen Voraussetzungen (geringe Gefahr von Flucht oder neuen Straftaten) ausgewählt worden sind.

⁴Das Propensity Score Matching ist ein methodisches Verfahren, das bei quasiexperimentellen Designs die Äquivalenz zwischen Behandlungs- und Kontrollgruppe sicherstellen soll. Dabei werden den behandelten Teilnehmern jeweils aufgrund einer statistischen Entscheidungsverfahren vergleichbare Personen aus einer größeren Grundgesamtheit von Nichtteilnehmern zugeordnet [20].

unbehandelt waren, sondern eben an anderen, wohl ebenso effektiven Maßnahmen teilnahmen.

Selbst diese ernüchternden Schlussfolgerungen erscheinen jedoch noch zu optimistisch, wenn man nach dem ITT-Ansatz die AGT-Abbrecher einbezieht (100 Personen, also 39,2 % der insgesamt 255 Personen, die ein AGT begonnen hatten). Die Wiederverurteilungsquote der Abbrecher betrug 35,0 %; dies ergibt für die Gesamtzahl der Teilnehmer (Absolventen plus Abbrecher) eine Quote von 28,2 %, die um etwa ein Sechstel über der Wiederverurteilungsquote der Nichtteilnehmer liegt und entweder auf einen insgesamt negativen Behandlungseffekt oder aber auf erhebliche Stigmatisierungseffekte für die Abbrecher schließen lässt.

Beispiel 5: „Trainingscamp Lothar Kannenberg“ als Alternative zum geschlossenen Jugendstrafvollzug

Im nördlichen Hessen gibt es seit 2004 das „Trainingscamp Lothar Kannenberg“ (seit 2014 „Trainingscamp Diemelsstadt“). Dies ist eine stationäre Einrichtung der Jugendhilfe, in der strafrechtlich auffällig gewordene Jugendliche sechs Monate lang v. a. durch sportliche Angebote und eine Variante der „konfrontativen Pädagogik“ betreut werden. Bei den Klienten handelt es sich um 14- bis 17-jährige männliche Jugendliche, die wegen ihrer Delinquenz durch Untersuchungshaft bedroht waren oder als richterliche Auflage an dem Programm teilnahmen und deren Betreuung vielfach schon in anderen Angeboten der Erziehungshilfe gescheitert war. In einer ersten Evaluationsuntersuchung der Universität Kassel [5] wurden die Rückfallraten der 154 Teilnehmer (93 Absolventen und 61 Abbrecher) den Vergleichsdaten aus der Legalbewährungsstatistik [11] gegenübergestellt. Die Ergebnisse fassen die Autoren dahingehend zusammen: „[es] deutet sich an, dass in Bezug auf den Aspekt der Legalbewährung ein nicht geringer Teil vom

Angebot des Trainingscamps profitieren kann, auch wenn Rückfälle durchaus häufig vorkommen“ ([5], S. 60).

Tatsächlich jedoch hängt die Bilanz der Maßnahme sehr stark erstens davon ab, welche Vergleichsgruppen man heranzieht, und zweitens, ob man die Abbrecher einbezieht (ITT) oder ausklammert (TR). Die entsprechenden Rückfallraten sind in Abb. 3 aufgeführt. Die Rückfallrate der Absolventen (59,1 %) liegt zwar deutlich unterhalb der Personen, die aus einer Jugendstrafe ohne Bewährung entlassen worden sind (78,0 %), sowie der Personen, die einen Jugendarrest verbüßt haben (70,0 %), ist aber fast identisch mit der Rückfallrate derjenigen, deren Jugendstrafe zur Bewährung ausgesetzt worden war (59,6 %). Allerdings lag auch hier die Rückfallrate der Abbrecher des Trainingscamps mit 75,4 % deutlich über der Rückfallrate derjenigen, die die Maßnahme regulär zu Ende geführt hatten („Absolventen“). Für alle 154 Jugendlichen, die die Maßnahme begonnen hatten, ergibt sich demnach eine Rückfallrate von 66,8 %, sodass sich die Erfolgsbilanz im Vergleich zu den formellen und insbesondere den stationären jugendgerichtlichen Sanktionen deutlich relativiert.

Beispiel 6: kognitive Trainings als ambulante Maßnahme in England

Die Effekte von zwei unterschiedlichen kognitiven Trainings („Think First“ und „Enhanced Thinking Skills“) im Rahmen von ambulanten Sanktionen wurden in England von [8] untersucht. Die Treatment-Gruppe bestand aus 173 männlichen Tätern, die die Weisung erhalten hatten, an dieser Maßnahme teilzunehmen. Diese Weisung wurde erteilt aufgrund eines erhöhten Risiko-Scores sowie festgestellter Defizite in kognitiven Fertigkeiten. Diesen Probanden wurden per Propensity Score Matching ebenfalls 173 Vergleichsprobanden, die aus einer größeren Grundgesamtheit ausgewählt worden waren, gegenübergestellt, die diese Weisung nicht hatten. Beim Matching wurden Risiko-Score, Lebensalter und Zahl der Vorstrafen berücksichtigt. Erfolgskriterium war die Wiederverurteilung, wobei der Beobachtungszeitraum zwischen 280 und 1454 Tagen (Mittelwert 760 Tage) variierte. Die Wiederverurteilungsquote war in der Kontrollgruppe (72,3 %) höher als in der Behandlungsgruppe (59,0 %). Jedoch war auch die Zeit in Freiheit bei den Unbehandelten länger. In der multivariaten Analyse (sequenzielle logistische Regression) ergab sich bei Einbeziehung dieser Faktoren nach dem ITT-Ansatz kein signifikanter Effekt der Behandlung; tatsächlich schnitt hier die Behandlungsgruppe sogar etwas schlechter ab als die Kontrollgruppe.

Jedoch war bei den Behandelten die Abbruchquote sehr hoch: Den 66 Absolventen der Maßnahmen („completers“) standen 69 „nonstarters“ (zur Behandlung vorgesehen, aber nicht angetreten) und 38 „noncompleters“ (Abbrecher)

gegenüber; die Ausfallquote betrug somit insgesamt 61,8 %. Die Rückfallraten der Nonstarters (72,5 %) und der Abbrecher (68,4 %) waren jeweils fast doppelt so hoch wie die der Absolventen (39,4 %). Die Autoren argumentieren nun, aufgrund des hohen Anteils von Ausfällen sei der ITT-Ansatz hier unpraktikabel, und durch die Einbeziehung der Ausfälle sei vermutlich ein positiver Effekt des Behandlungsabschlusses durch einen negativen Abbrucheffect „maskiert“ worden. Um diese Effekte zu separieren, verglichen sie nun im Rahmen eines TR-Ansatzes die Rückfallquoten der drei Gruppen (Absolventen, Nonstarters und Abbrecher) mit den Rückfallraten der drei Sektionen aus der Kontrollgruppe, die sich jeweils durch Matching ergaben. Das bedeutet, den 66 Absolventen wurden ihre gematchten „Zwillinge“ aus der ursprünglichen Kontrollgruppe zugeordnet; ebenso den Abbrechern und den Nonstarters. Im Ergebnis sehen sie „moderate Effekte des Programm-Abschlusses“, da die Absolventen nun wieder besser abschnitten als die gematchten Mitglieder der Kontrollgruppe.

Dieser scheinbare Beleg für die Existenz eines Abschluss-effekts ist jedoch nicht überzeugend. Denn erstens schnitten zugleich die Abbrecher und Nonstarters schlechter ab als ihre jeweiligen Kontrollgruppen, sodass ein Vorteil der Behandlung insgesamt weiterhin nicht erkennbar wird. Es ist auch weiterhin nicht auszuschließen, dass der vorgebliche Abschlusseffekt nur in einem Ausfalleffekt besteht, d. h., dass der selektive Ausfall dazu führte, dass Personen mit günstigerem Risikoprofil die Maßnahme abschlossen. Zweitens ist die von den Autoren [8] gewählte Vorgehensweise der Bildung von Untergruppen der Kontrollgruppe nicht geeignet, die Verletzungen der Äquivalenz durch Ausfalleffekte zu reparieren. Da die beiden durch das Matching erzeugten Gruppen hinsichtlich der Faktorkombination, die den Abbruch steuert, nicht exakt dieselbe Rangreihe aufweisen müssen, entstehen auf diese Weise keine jeweils äquivalenten Vergleichsgruppen. Der dafür verantwortliche Effekt (Regression zur Mitte) lässt sich an folgendem Gedankenexperiment verdeutlichen: Angenommen, für ein sportliches Trainingsprogramm werden 100 Brüderpaare ausgewählt, von denen jeweils ein Bruder der Trainingsgruppe und einer der Kontrollgruppe zugelost wird. Angenommen weiter, in dem Training gibt es 40 % Abbrecher, wobei ein wesentlicher Faktor für den Abbruch darin besteht, dass für unsportliche Teilnehmer das Training unangenehm ist. Wenn man nun den 60 Absolventen des Trainings nur ihre 60 gematchten Partner aus der Kontrollgruppe zuordnet, werden diese vermutlich nicht alle genauso sportlich sein wie die Absolventen der Trainingsgruppe, und zwar deshalb, weil zwischen Brüdern das Merkmal Sportlichkeit nicht hochgradig konkordant ist. Falls die 60 Sportlichsten der Trainingsgruppe das Training bis zum Ende durchstehen, werden unter ihren 60 Brüdern auch einige weniger Sportliche sein, d. h., das Merkmale „Sportlichkeit“, auch wenn es in den

beiden ursprünglich vorgegebenen Bedingungen nicht verschieden war, wird durch den Ausfalleffekt in der Gruppe der Absolventen einen höheren Mittelwert aufweisen als in dem entsprechenden Sektor der Kontrollgruppe. Das individuelle Matching post hoc (nach erfolgtem Abbruch) gewährleistet also in der Regel nicht, dass die so entstehenden Teilkontrollgruppen tatsächlich gut vergleichbar sind. Eine bessere Leistung der Trainingsgruppe in einem Sporttest kann also trotz dieser Vorkehrung immer noch auf einen selektiven Ausfalleffekt zurückzuführen sein, sie muss nicht zwangsläufig auf den Abschluss des Trainings zurückgehen.

Fazit und Schlussfolgerungen

Die aufgeführten Beispiele belegen die theoretischen Überlegungen, denen zufolge nur der ITT-Ansatz geeignet ist, die Wirksamkeit einer Behandlung zu evaluieren, weil nur er es ermöglicht, methodische Verzerrungen in den Daten aufgrund selektiver Abbrüche (Ausfalleffekt) weitgehend auszuschalten. Dies gilt für die Straftäterbehandlung gleichermaßen wie in der medizinischen oder psychotherapeutischen Behandlungsforschung. Die in Veröffentlichungen mitgeteilten Ergebnissen deuten manchmal, wenn nur die „Absolventen“ mit einer Kontrollgruppe verglichen werden, auf einen gewissen Behandlungseffekt hin, der jedoch oftmals verschwindet (und sich in einigen Fällen ins Negative dreht), wenn die Abbrecher mitberücksichtigt werden. Nur Studien, die einen ITT-Ansatz verwenden, sind geeignet, die Wirksamkeit einer Behandlung zu belegen. Und nur Maßnahmen, die per saldo wirksam sind (in dem Sinne, dass die Bereitstellung dieser Behandlung zu einer Verbesserung des Ergebnisses führt, im Fall der Straftäterbehandlung also zu einer Verhinderung von Rückfällen), können sowohl politischen Entscheidern als auch einzelnen potenziellen Teilnehmern auf rationaler Basis empfohlen werden.

Der TR-Ansatz (Vergleich der Behandlungsabsolventen mit der Kontrollgruppe) ist eher dazu geeignet, die Leistungsfähigkeit einer Behandlung („efficacy“) abzuschätzen, d. h. das Ausmaß der Wirkungen, die bei kooperativen Personen unter günstigen Bedingungen möglich sind. Diese Leistungsfähigkeit ist jedoch von der Wirksamkeit („effectiveness“) zu unterscheiden. Eine Behandlung, die nur in der Teilgruppe der Absolventen die erwünschten Wirkungen zustande bringt, deren Wirksamkeit insgesamt jedoch aufgrund negativer Effekte bei den Abbrechern nicht zu belegen ist, kann nicht als bewährte Methode gelten. Studien mit einem TR-Ansatz begünstigen irreführende Schlussfolgerungen, wenn man aus ihnen Belege für die Wirksamkeit herauslesen will.

Bei der Interpretation einzelner Studien ist es oft allein aufgrund der mitgeteilten Befunde kaum möglich zu entscheiden, ob das schlechtere Abschneiden der Abbrecher

auf selektive Ausfalleffekte (Personen mit schlechteren Risiken brechen häufiger ab), auf inhaltliche Abbrucheffekte (Stigmatisierung und Motivationsverlust) oder auf negative Behandlungseffekte in Teilgruppen zurückzuführen ist. Während jedoch Selektionseffekte vielfach belegt sind (die bekannten Risikofaktoren sind auch häufig Determinanten des Abbruchs; [17, 3]), gibt es wenig empirische Evidenz für die rückfall erhöhenden Effekte von Fremd- oder Selbst-Stigmatisierung nach Behandlungsabbruch. Häufig zu lesende Empfehlungen, die Abbruchraten von Behandlungsprogrammen zu vermindern, um dadurch deren Wirksamkeit zu steigern, beruhen offensichtlich auf der ungeprüften Annahme, die hohe Rückfallrate der Abbrecher sei allein durch den Abbruch bedingt, und von diesen Personen würde ein geringerer Anteil rückfällig werden, wenn man sie an der Maßnahme bis zum Ende hätte teilnehmen lassen. Das mag vielleicht so sein, ist aber selten belegbar, und die Empfehlung könnte durchaus auch kontraproduktiv sein: Ungeeignete Teilnehmer in einer Maßnahme zu belassen, von der sie nicht profitieren können, kann bedeuten, dass Ressourcen fehlalloziert werden, und die ungeeigneten, unmotivierten oder undisziplinierten Teilnehmer können insbesondere in Gruppenmaßnahmen den therapeutischen Erfolg der anderen Teilnehmer gefährden.

Negative Behandlungseffekte sind in der Medizin an der Tagesordnung, und entsprechende Angaben sind wesentlicher Bestandteil eines jeden Beipackzettels für Medikamente. In der Psychotherapie werden sie für einige Therapieverfahren beschrieben [15]. Für die Straftäterbehandlung wurden sie bisher nur in einigen Aspekten (z. B. rückfall erhöhende Wirkung unstrukturierter Behandlungsmaßnahmen bei Psychopathen [19]) diskutiert. Es erscheint jedoch nicht unplausibel, dass die in vielen Studien ausbleibenden Effekte von Behandlungen manchmal darauf zurückzuführen sind, dass positive Effekte bei einem Ausschnitt der Zielgruppe durch negative Effekte bei einem anderen Ausschnitt kompensiert werden.

Die bisherigen Überlegungen beziehen sich primär auf randomisierte experimentelle Studien, bei denen die Behandlungsabsicht durch die Untersuchenden vorgegeben wird. Dabei gilt für Personen, die für die Behandlung vorgesehen waren, diese aber nicht aufgenommen haben (Nonstarters) im Prinzip nichts anderes als für Abbrecher, die erst im Verlauf der Behandlung ausgestiegen sind (Drop-outs). Diese Argumente sind aber auch auf quasiexperimentelle Studien zu übertragen. Wenn man weiß, dass Behandlungsabbrüche in einer randomisierten kontrollierten Studie i. Allg. einen Störeffekt darstellen, muss man auch für vorgefundene Gruppen von Behandelten und Unbehandelten, deren Äquivalenz zunächst angenommen wird, erwarten, dass selektiver Abbruch diese Äquivalenz gefährdet und sich zugunsten nur scheinbarer Behandlungseffekte auswirkt. Diesem Störeinfluss muss also auch hier durch einen

strengen ITT-Ansatz entgegengewirkt werden. Problematisch erscheint es, wenn als Vergleichsgruppe vorwiegend Behandlungsverweigerer verwendet werden.

In Hinsicht auf die oben postulierten unterschiedlichen Effekte, mit denen bei Behandlungsstudien zu rechnen ist, liegen die folgenden Empfehlungen nahe:

- Hinsichtlich der *intendierten positiven Behandlungseffekte* kommt es darauf an, geeignete Programme zu entwickeln und zu optimieren. Dabei ist auch zu berücksichtigen, dass nach dem Risk-Need-Responsivity-Ansatz (RNR) [1, 4] Behandlung v. a. dann rückfallpräventiv wirkt, wenn sie an relevanten dynamischen Risikofaktoren ansetzt; das betrifft die Frage nach der differenziellen Indikation: Für wen ist das Programm geeignet? Unter welchen Voraussetzungen ist es angezeigt?
- Die Wirkung einer Behandlung kann sich natürlich nur bei den Personen einstellen, die an ihr teilnehmen, bis zum Ende oder jedenfalls in einem gewissen Umfang. *Abschluss- und Dosis*effekte sind somit die Voraussetzung dafür, dass insgesamt ein positiver Gesamteffekt beobachtet werden kann. Bei maximalen Abbruchraten können sich diese Effekte nicht einstellen und damit auch keine positive Gesamtbilanz der Maßnahme. Deshalb kommt es darauf an, die Zielgruppen zu bestimmen, bei denen die Maßnahme wirkt und somit indiziert ist. Aus diesem Grund ist die Wirksamkeit einer Behandlung auch keine feste und absolute Größe, sondern stets auf eine spezifische Zielgruppe zu beziehen.
- Die Frage nach der Indikation ist besonders dann zu stellen, wenn man annimmt, dass intensive Interventionen auch *negative Auswirkungen* haben können, insbesondere dann, wenn sie bei Personen angewendet werden, für die sie nicht geeignet sind. Die Forschung muss klären, bei welchen Personen evtl. sogar rückfall erhöhende (Neben-)Wirkungen zu befürchten sind; für diese ist die Behandlung dann als kontraindiziert anzusehen, und alternative Maßnahmen sind zu entwickeln.
- *Abbrucheffekte in Form von Stigmatisierungseffekten oder Entmutigung* bei Abbrechen sind, falls sie auftreten, durch das Verhalten des sozialen Umfelds oder der Institutionen bedingt. Falls die Abbrüche nicht vermeidbar sind, käme es dann darauf an, die Folgen abzufedern, durch das Verhindern von Ausgrenzung und Diskriminierung (Verweigerung von Lockerungen), und alternative Angebote zu unterbreiten.
- *Selektive Ausfalleffekte*, die bei Behandlungsabbrüchen dazu führen, dass sich Personen mit günstigen und ungünstigen prognostischen Aussichten trennen, lassen sich nicht nur als methodische Störeffekte betrachten. Sie können zum einen dafür nützlich sein, die für eine Behandlungsmaßnahme geeignete Zielgruppe zu identifizieren. Zum anderen könnten für Personen, die

eine erhöhte Neigung zum Abbrechen erkennen lassen, stützende und motivierende Hilfen angeboten werden. Des Weiteren ist jeweils zu prüfen, von welchen Personenmerkmalen oder Umständen erhöhte Abbruchraten in Teilgruppen abhängen; dies kann Anlass geben, das Behandlungsprogramm im Sinne des Ansprechbarkeitsprinzips [1] so zu modifizieren, dass es besser zu den kognitiven und motivationalen Voraussetzungen dieser Zielgruppen passt, oder für diese Gruppen alternative Angebote zu entwickeln.

Einhaltung ethischer Richtlinien

Interessenkonflikt J. Endres, M.M. Breuer und M. Stemmler geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

Literatur

1. Andrews DA, Bonta J (2010) The psychology of criminal conduct. 5. Aufl. Anderson, New Providence
2. Bortz J, Döhning N (2006) Forschungsmethoden und Evaluation, 4. Aufl. Springer, Berlin
3. Endres J (2014) Determinanten der Behandlungsteilnahme und des Behandlungsabbruchs bei inhaftierten Sexualstraftätern. Forum Strafvollzug 63:237–243
4. Endres J, Schwanengel MF (2015) Straftäterbehandlung. Bewährungshilfe 62:293–319
5. Galuske M, Böhle A (2010) Evaluation des Trainingscamps Lothar Kannenberg. Z Jugendkriminalrecht Jugendhilfe 21:52–61
6. Greenhalgh T (2010) How to read a paper. The basics of evidence-based medicine. Wiley-Blackwell, Chichester
7. Gupta SK (2011) Intention-to-treat concept: a review. Perspect Clin Res 2:109–112
8. Hatcher RM, McGuire J, Bilby CAL, Palmer EJ, Hollin CR (2012) Methodological considerations in the evaluation of offender interventions: the problem of attrition. Int J Offender Ther Comp Criminol 56:447–464
9. Hirtenlehner H, Hiebinger I (2013) Rückfallergebnisse eines gruppenorientierten Antigewalttrainings in der Bewährungshilfe – Befunde aus Österreich. Z Jugendkriminalrecht Jugendhilfe 24:57–64
10. Holling H, Schmitz B (Hrsg) (2010) Handbuch Statistik, Methoden und Evaluation. Hogrefe, Göttingen
11. Jehle JM, Heinz W, Sutterer P (2003) Legalbewährung nach strafrechtlichen Sanktionen: Eine kommentierte Rückfallstatistik. Forum Verlag Godesberg, Mönchengladbach
12. Jehle JM, Albrecht HJ, Hohmann-Fricke S, Tetal C (2013) Legalbewährung nach strafrechtlichen Sanktionen: Eine bundesweite Rückfalluntersuchung 2007 bis 2010 und 2004 bis 2010. Forum Verlag Godesberg, Mönchengladbach
13. Kendall PC, Holmbeck G, Verduin T (2004) Methodology, design, and evaluation in psychotherapy research. In: Lambert MJ (Hrsg) Bergin and Garfield's handbook of psychotherapy and behavior change. Wiley, New York, S 16–43
14. Kwong DLW, Darby S, McGale P et al (2011) Effect of radiotherapy after breast-conserving surgery on 10-year-recurrence and 15-

- year breast-cancer death: meta-analysis of individual patient data for 10801 women in 17 randomised trials. *Lancet* 378:1707–1716
15. Lilienfeld SO (2007) Psychological treatments that cause harm. *Perspect Psychol Sci* 2:53–70
 16. Marques JK, Wiederanders M, Day DM, Nelson C, von Ommeren A (2005) Effects of a relapse prevention program on sexual recidivism: final results from California's sex offender treatment and evaluation project (SOTEP). *Sex Abuse* 17:79–107
 17. Olver ME, Stockdale KC, Wormith JS (2011) A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *J Consult Clin Psychol* 79:6–21
 18. Ortmann R (2002) Sozialtherapie im Strafvollzug: Eine experimentelle Längsschnittstudie zu den Wirkungen von Strafvollzugsmaßnahmen auf Legal- und Sozialbewährung. Max-Planck-Institut für ausländisches und internationales Strafrecht, Freiburg
 19. Rice ME, Harris GT, Cormier CA (1992) An evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law Hum Behav* 16:399–412
 20. Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 39:33–38
 21. Sherman LW (2003) Misleading evidence and evidence-led policy: making social science more experimental. *Ann Am Acad Polit Soc Sci* 589:6–19
 22. Spokas ME, Rodebaugh TL, Heimberg RG (2005) Treatment outcome research. In: McKay D (Hrsg) *Handbook of research methods in abnormal and clinical psychology*. Sage, Thousand Oaks, S 513–530
 23. Stelly W (2014) Die bessere Alternative zum geschlossenen Regelvollzug? 10 Jahre Jugendstrafvollzug in freien Formen. *Z Jugendkriminalrecht Jugendhilfe* 25:257–262
 24. Stemmler M (2015) Methodische Zugänge der Intervention und Evaluation. In: Hurrelmann K, Bauer U, Grundmann M, Walper S (Hrsg) *Handbuch Sozialisationsforschung*, 8. Aufl. Beltz-Juventa, Weinheim, S 348–362